DOCUMENT RESUME

ED 266 441 CS 008 343

AUTHOR Davison, Alice

TITLE Readability--The Situation Today. Technical Report No. 359.
INSTITUTION Bolt, Beranek and Newman, Inc., Cambridge, Mass.; Illinois Univ., Urbana. Center for the Study of Reading.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Dec 85
CONTRACT 400-81-0030
NOTE 48p.
PUB TYPE Information Analyses (070)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Difficulty Level; Language Processing; Measurement Techniques; *Readability; *Readability Formulas; Reading Comprehension; Reading Improvement; *Reading Research

ABSTRACT

A survey of the current research on readability formulas is presented in this paper, which distinguishes this research from research on the more general questions that surround formulas: What features of a text, particularly the language it is written in, make the text easy or difficult to read? and, What will predict that readers with particular levels of skills can read a particular text? It then discusses the features of readability formulas that have made them useful and appealing. The next section reviews research that illustrates the inappropriateness or lack of usefulness of readability formulas in cases involving (1) coping with jury instructions and government relations and forms, (2) matching children's books with readers of the right age and level of ability, and (3) choosing school materials of appropriate levels in languages very unlike English. The last sections explore research on readability formulas, addressing the issue of what constitutes complexity of language, and raise the question of which measures of language processing are most sensitive to features of language, and why. Four pages of references are included. (EL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED266441

CENTER FOR THE STUDY OF READING

Technical Report No. 359

READABILITY--THE SITUATION TODAY

Alice Davison
University of Illinois at Urbana-Champaign

December 1985

University of Illinois
at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

Bolt Beranek and Newman Inc.
10 Moulton Street
Cambridge, Massachusetts 02238

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

The work upon which this publication is based was performed pursuant to Contract No. 400-81-0030 of the National Institute of Education. It does not, however, necessarily reflect the views of this agency. To appear in Beverley L. Zakaluk and S. Jay Samuels, Readability: Its past, present and future, International Reading Association.

CS 008 343

Abstract

This paper distinguishes research on readability formulas from research on the more general questions which surround formulas: what features of a text, particularly the language it is written in, make the text easy or difficult to read? It discusses a number of different approaches to characterizing texts, which do not make use of formulas. Finally, the question is raised as to what measures of language processing are most sensitive to features of language, and why this is the case.

Readability--The Situation Today

This paper presents a survey of current research on readability, taking the term in a much more general sense than it is usually taken. The main point to be considered is that there are wider and narrower senses in which the term readability can be taken. In the narrower sense, it refers to the development and use of readability formulas and related objective methods which use a small number of measures of variables such as average number of words, syllables, etc., in a sentence or text. For a series of excellent surveys and discussion of work on readability formulas, including their successes and failures, there is no better source than the book and articles by G. Klare (Klare, 1963, 1974-75, 1984). But readability formulas were first created to answer a number of very broad questions--what makes a text difficult to read? What will predict that readers with particular levels of skills can read a particular text? (Here, the word text is used in its technical sense as sequence of connected sentences.) These questions remain largely unanswered even today, if we think in terms of a model of reading comprehension applied to linguistic features of the text. There has been much interesting and productive research on features of texts, such as general content and overall organization, in relation to readers' knowledge and ability to make sense of information. But very little is understood about how the

structure of sentences and the nature of the words used might affect comprehension of a text.

The successes of formulas have been of a statistical nature. For a large number of readers with varying abilities, and for large numbers of texts with varying sentence and word lengths, formulas can be used to make fairly successful predictions. But for more specific cases, they become less and less sensitive to special features of texts and readers. One particular question is often asked: Can this student, or this group of people, read this text? If not, why not? What can be done to improve the chances that certain readers will comprehend a certain text?

In addition, there is a problem of general theoretical interest. Readability formulas measure averages for length of sentence, and length or complexity familiarity of words, which can vary in different parts of a longer text. These measures are supposed to reflect complexity of language in some way, which, in some intuitive way, creates some barriers to comprehension. The nature of the barrier, or at least one type of obstacle to comprehension, is plausibly described as some sort of overload on the ability of the reader to process a certain quantity of linguistic information in a single short interval. But we know very little about what is affected and how.

Typical discussions of readability can be understood as interest in readability formulas, with the specific issues appropriate to these statistically based, objective predictive

devices, or, on the other hand, as a set of more general issues, many of which are completely independent of readability formulas. The central issue is the question of what features of a text contribute to difficulty in comprehension of its content. The question of difficulty may include linguistic variables, such as sentence structure and complexity of words or the information conveyed by the words. But it also includes the abilities of the reader, as well as the reader's background knowledge and perception of the situation in which reading a particular text is taking place. A great deal of research has been done on readability in the first sense, and is still going on in very much the same way that it has been going on since the formulas came into use. Not so much has been done in psychology, education or linguistics to provide answers within a rigorous model of how language is processed and comprehended in various situations. In the rest of this paper, a survey of various kinds of research which are being done, and which promise a way of approaching more satisfactory answers is presented. It is here proposed that only technical refinements can be made in research on readability formulas, and without research of the second kind, focussing on the fundamental questions of how language is read and understood, we will not make much progress in understanding readability or in more effectively matching texts and readers.

Current Research

The two features of readability formulas which have made them useful and constitute their appeal can be summarized as follows:

(1) they measure features of language in an objective way, with statistical accuracy in their predictions of levels of comprehension.

(2) as a sampling procedure, taking average values for small parts of larger texts, they reduce the task of assessing difficulty, and the calculations can be done without special training or equipment.

For the moment we ignore potential challenges to these assertions. Much current research, as noted in Klare (1984) has been devoted to these two issues. That is, research has been concentrated on the statistical features of formulas. Norms have been recalculated for the McCall-Crabbs reading passages which serve as the criterion for the predictions of formulas. Certain formulas have been revised to reflect the performance of contemporary student populations, and others have been created to make predictions for adult readers reading technical materials. It is likely that general formulas will continue to be adapted for adult readers and non-school materials. One of the strongest current demands placed on formulas is for predictions for adult readers, especially those with poor reading skills, who must

read, at a high level of comprehension, technical or other demanding reading material.

Research also continues to be done on the measures or predictors of readability, the features of the language in a text for which objective calculations are made. Because of the growing use of computers in finding and testing statistical correlations, and in integrating enormous amounts of information, it is possible to explore in much greater detail than before all the possible ways that readability levels can be calculated, and to find more and more specific features of text (letters per word, number of coordinating conjunctions, number of anaphoric words, etc.) which serve as predictors of difficulty. It is also possible to avoid a problem of sampling by taking many more samples of text at regular intervals, or even to calculate formula values for entire texts.

The use of large amounts of data with the help of computers has helped to overcome some of the criticisms which have been made in the last few years, that older formulas were out of date, and that word lists of familiar words and the McCall-Crabbs reading passages did not necessarily reflect reading skills today of the student population; that they automatically make accurate predictions for adults and for the kind of technical materials which adults are called upon to read, such as instructions for forms, maintenance manuals, tax forms, etc. The ability of computers to deal efficiently with large amounts of data has also

overcome some of the objections to formulas based on sampling of passages from texts. Formulas in themselves often don't specify a sampling procedure which contributes to accurate predictions, and they certainly don't guarantee that a correct sampling is performed.

The use of computers helps to ensure that readability formulas make accurate statistical predictions to the extent that they are capable of doing so. But this trend, with all its advantages, introduces a certain contradiction in the use of formulas. If computation of readability levels requires the use of computers and the skills necessary to apply computers for this purpose, then the use of readability formulas is no longer in the hands of the average user, though this situation will probably change a little with the growing availability and use of microcomputers. Hence very detailed and accurate use of formulas is not always within the reach of the ordinary user of the formulas.

Another very striking trend in research on readability has concentrated on making existing formulas easier to apply than before. In some cases, this involves more efficient hand-counting of the linguistic variables, in others, it means more efficient calculation of the factors in the formulas--this is facilitated recently by the availability of small calculators as well. The Raygor Readability Estimator (Raygor, 1977) is a splendid example of both aspects of simplification. Instead of

counting all of the syllables in a 100 word sample, one counts only those words with more than 6 letters, a number of letters which can be determined by eye in most cases, rather than by actual counting of letters. The number of words of six letters or more is entered on a slide-rule like calculator, and the grade level is then read off a scale in relation to the number of sentence breaks in the sample. The small size and compactness of the calculator and the simplification of the counting procedure in fact make it very easy to use it in conjunction with a text of any length. The calculator itself is large enough to contain a printed warning about what kind of sampling procedure to use, what kind of text to apply it to and specifically which kinds of text not to apply it to, and finally what degree of accuracy to expect. If the user reads this set of instructions, then formulas will be applied with a reasonable sampling to the right kind of text, and the result will be a prediction of an approximate readability level.

There are two strong trends, then, in current readability research. One is towards greater statistical accuracy and more comprehensive measurement of text variables, achieved through large manipulations of data, and the other is toward greater convenience for the average computationally unskilled user, with some loss in fineness of detail or statistical accuracy. Clearly these trends are in conflict, and one might ask if one and the same formula can really be asked to serve two such differing

purposes. One might also speculate that some different directions might be taken in the two areas of greater accuracy and user convenience.

Some of the research in a new direction might deal with new aspects of texts. As Klare notes (1984, p. 685), formulas are not sensitive to most of the important features which seem to affect how well a reader will comprehend the text. These include content, style, format and organization, and of these formulas measure only style. It might be argued that they do not measure style either, except in the narrowest sense of sentence length and word complexity. Other features of style which are of a more 'literary' quality include the use of sentence structure and word choice to convey aspects of meaning in addition to the literal content of the text. But in any case, formulas are not sensitive to the motivations of the reader, the purpose for reading and the amount of background knowledge which the reader already has about the subject matter in the text.

It might be possible to reduce these factors to formula-like variables, and to do statistical correlations for them, as with the other variables used. Of course, many of the linguistic factors are both difficult to identify without careful prior analysis of the text, and also infrequent in statistical terms. Other factors such as text organization are difficult to reduce to objectively definable units, particularly since we know very little about how discourses are really structured. Finally, we

know very little about how factors such as text organization and syntactic structures interrelate, if in fact they do. It appears that the extension of formulas to cover other variables would be useful and effective only if we had some well-founded hypotheses about how they affect comprehension. Statistical correlations with comprehension might be obtained by a trial and error procedure, but even if the results were interesting it is unlikely that they would be as informative about the process of comprehension as direct observation.

Current refinements of readability formulas may make the approach as effective as it will ever be for predictions about large aggregates of texts and readers. But we will not begin to understand what makes a text complex and under what circumstances, unless we look directly at aspects of texts, readers and situations. To do this we need to be concerned with understanding in a more general way how language is comprehended, and how skills are acquired in interpreting linguistic structures. In other words, the real questions of readability are questions of educational and cognitive psychology, linguistics and cognitive science, in general.

Current Approaches to Research on Readability not Involving Formulas

In this section, we present some recent research which illustrates possible alternative approaches to dealing with the complexity of texts and of the language in which they are

written. This survey is not meant to be exhaustive; for additional references and discussion, see Klare (1984, p. 701ff). Though it is not exhaustive, the survey includes a diverse group of examples, including ones which lie outside of the topics usually discussed in connection with formulas, in order to underline the fact that there is no single best approach to the great variety of problems of text difficulty. Rather, each specific situation needs to be approached in the terms appropriate to its set of internal features--what readers are involved, what their purpose in reading is, and the nature of the texts and language involved. Of course it is to be hoped that when we have gained more understanding than we have at present about how language and information are understood and remembered, then perhaps some unifying principles will emerge.

The cases discussed below all involve the need to know what makes a text linguistically complex, and how to make it less so, or else how to match readers of different levels with texts within their ability. Two involve adult readers coping with technical materials: jury instructions and government regulations and forms. Another concerns the match of children's books with readers of the right age and level of ability, outside of the context of school reading. Others are samples of projects being done in many societies, involving languages very much unlike English, perhaps with no tradition of writing, where school materials of appropriate levels must be chosen or written.

In all of these cases, readability formulas are inappropriate or not useful, and other resources must be chosen from the ones which are available.

Jury instructions. Robert and Veda Charrow (1979) studied how well adults comprehend the legal definitions given to members of juries, and compared the level of comprehension for the usual form in which the instructions are given with a form revised by changing the specific linguistic factors which were correlated with poor comprehension. Jury instructions are definitions of principles of law, such as what constitutes contributory negligence. These standard definitions, composed by lawyers, are read to the members of the jury before they begin their deliberations. Their decision is to be related to these points of law--that is, if the defendant is guilty of contributory negligence, in this particular definition. There is both anecdotal and systematic evidence that most jurors, even those with education beyond high school, do not understand these definitions very well, though the more education a juror has, the better the instructions are understood. But clearly it is desirable that the average juror should be able to understand the principles which guide his or her decision.

In the first part of the study, the sources of difficulty were located. A test of recall, which reflected comprehension, showed that difficulties of comprehension were associated with specific semantic and syntactic characteristics of the text.

These included double negations, parenthetical elements placed far away from the material they were related to, multiple subordinate clauses, and deleted elements. The revisions were made by substituting more explicit, less complex but equivalent sentence structures, with the main goal in mind of presenting the original content in a clear and perspicuous way. In general, the difficulty levels of the originals, as measured by readability formulas, did not change in the revisions.

In the second part of the study, two groups of prospective jurors were asked to listen to the same jury instructions, half in their original form, and half in their revised form. Because each group saw some original and some revised instructions, it was possible to compare performance for the two forms of each instruction. The revised forms were comprehended better than the original forms, by a significant amount. The increase in comprehension was about 40% over the level of comprehension found in the original form.

From the point of view of the real world problem of making sure that jurors are adequately informed about the decision they are asked to make, the changes reported by the Charrows are not enormous. In some cases, the original level of comprehension was 25%, but the improvement reached only 42%; we would rather have all or nearly all the jurors understand the instructions completely. But these results are still very interesting and important for two reasons. First, the increases in comprehension

were unrelated to readability formulas. The readability levels for the revisions would not have predicted the observed gains in comprehension, and in fact did predict increases for four instructions for which no significant results were seen. Second, the revisions were guided by features not connected with readability formulas; the sentences were not shortened and the long words were not replaced by shorter or more frequent ones. The investigators attempted to diagnose the possible difficulties in the text by looking at both the content and the form of the text. The increase in comprehension appears to be caused by changes in the outward form of the text, the clarifications in the syntax and organization, in spite of the fact that the content, which was complex, remained the same. One of the most interesting and useful features of Charrow and Charrow (1979) is the detailed discussion of each instruction, its particular difficulties and how they were resolved.

Government regulations and forms. It is widely perceived that government forms and regulations are very difficult for lay people to read and understand correctly, particularly those with little education and no access to expert help. One trend in the movement toward simplification has been to apply readability formulas, to shorten sentences and simplify words, though with no evidence that the predictive power of formulas extends to what are very special and fragmentary texts of this kind (for discussion see Holland, 1981. Another and unfortunately less

popular trend involves making use of information from potential readers which can be used to get insights in how people go about understanding unfamiliar and abstract texts of this kind, and how revisions can make better use of the resources which readers bring to the task of reading.

In a revision of Medicaid forms, Redish and others (reported in Holland, 1981) found that the users of the form were not clear about the meaning of some of the words and phrases used. They noted, however, that caseworkers who used the original difficult form had evolved ways of paraphrasing the difficult parts and of giving specific explanations for questions. Some of these explanations were incorporated into the revision. Flower, Hayes, and Swarts (1980) found that people attempting to read complex and abstract material such as government regulations do not concentrate so much on deciphering the long or complex sentences and hard words. What they do, as a strategy for understanding, is to translate abstract statements into specific instances, which have the form of a series of related events, or a scenario. In a scenario the actors have particular goals and react to specific circumstances. Information expressed in this form, as a sequence of related events with identifiable cause and effect relations, seems to be clearer than the equivalent information summarized in condensed and abstract terms. People may also typically not realize what connections there are between items in a form, since they are not familiar with forms and the

purposes they are used for. The group who revised the Medicaid form tried to help the applicants to see that the form represented a coherent whole, with relations among the questions, by presenting the form as a kind of letter. Most people do know what kind of text a letter is, and they expect there to be connections among the parts of a letter.

Some increases in comprehension are often found with reader-based revisions like these (cf. Holland, 1981), but in some cases there are no observable effects. Walmsley, Scott, and Lehrer (1981) compared original and revised forms of health-related documents which were read by elderly people who answered questions about the content. The revisions were done either to reduce the readability formula levels of difficulty of the originals, or to correct for difficulties in the text which skilled writers could perceive and change in the originals. Only for the longest of the documents were any differences found in level of success in answering comprehension questions. The revision made by skilled writers for this one document showed gains for both good and poor readers, while the revisions done in accordance with formulas showed no overall gain in comprehension, and even some loss. But readers showed a preference for all four documents in the revision done by skilled writers. So even if revisions done with the readers and the content as the primary factors produce a gain of 0% to 10% in comprehension, it might be worthwhile to pursue this kind of revision because the results

seem to make the task of reading this kind of material less onerous and less unpleasant.

With adult readers and rather specialized texts to be read, it might be expected that readability formulas lose a lot of their predictive power, since the statistical strength of formulas is in large aggregates of different texts and different levels of ability (Rodriguez & Hansen, 1975). One response to this is to evolve very specialized formulas for a particular class of readers and a class of texts with particular content. But while this approach might restore some of the statistical predictive power of a formula, it remains a superficial way of treating texts and readers. Alternatively one could devote time and effort to understanding how readers understand texts and what particular difficulties they encounter. A formula makes certain predictions, which may or may not hold in a specific instance, and there is no way of finding out why a given reader did or did not cope with a text. The studies just surveyed were done in order to define features of text which could be made easier to understand for the audience in question, and in particular to find out what resources the readers could use even if they were not highly skilled at reading.

Stories in Children's Reading Lessons

The subject of the research discussed here is quite familiar in the context of readability formulas. Formulas are often applied to the stories in children's reading textbooks to

determine their relative difficulty, and revisions are often made in the text of these stories to improve the readability levels assigned them (but see Green & Laff, to appear, for evidence of the effectiveness of revisions). Beck, McKean, Omanson, and Pople (1984) compared two versions of two stories of approximately second or third grade levels for how much of the story children were able to recall and how well they were able to answer comprehension questions. What is of particular interest here is how the revised versions of the stories were created. A close analysis of the two stories was made to find possible sources of difficulty in the original texts. The revised version involved changes in these features of the text, changes which were designed to correct for the difficulties.

Most of these possible sources of confusion stemmed from ways that the content of the story was expressed, either in linguistic factors or what was expressed versus what was implied. The linguistic factors included unclear reference to things in the text, ambiguous reference to antecedents and inexplicit or ambiguous temporal and causal relations. Problems with content included distractions in the text caused by irrelevant details, and unexpressed important details which were meant to be inferred in the original. Note that these factors are ones which a skilled writer or editor would pick out as flaws in a text which was supposed to be clear and felicitous--that is, to contain in the surface expression of the text, information which would help

the reader to understand the content. This information is especially important for younger, less skilled readers with imperfect background knowledge. Readability formulas are not sensitive to these text factors.

For both skilled and less skilled readers in third grade, the revised versions were understood better. There was greater recall of the stories and greater success in answering comprehension questions. As in other studies of this nature, the gains were not tremendous and performance overall was not impressively good. The percentage of correct answers to comprehension questions was 60% for the original version and 66% for the revised version. The level of success for less skilled readers increased as much as the scores of skilled readers, for the revised versions. But in a related study (Omanson, Beck, Voss, & McKeown, 1984) the nature of the form of the stories was not changed but the reading lesson was revised, so that questions about segments of the stories were made more explicit and more closely related to the text being read. The revised reading lesson questions led to recall of much more of the central parts of the stories. In the unrevised condition, the parts of the stories which 50% or more of the children recalled were short, fragmentary and omitted the points on which the stories hinged. For the revised questions, the parts of the stories recalled included not only the main characters but also more of the sequence of important events. Again, increased comprehension is

achieved without manipulating the text in ways which would change the readability levels assigned to the texts. The increase is significant, although it may not seem enormous, and it does not approach perfection.

Nevertheless, these approaches to increasing comprehension of a basal reader have a great deal of significance as interventions which are totally independent of readability formulas. The text elements which are affected are not those which could be picked out by a readability formula or even a taxonomy of difficult constructions. The changes made in the text do not alter the readability levels which would predict comprehension. What is most important, however, is that these interventions go directly to the central issues, reading a well-formed text and learning to pay attention to information in a text. It is more defensible to make sure that children use their efforts to read texts which are not basically ill-formed and flawed, ones which have in them what children are learning to pay attention to and understand.

Children's Literature

Books published for children to read, or have read to them, outside of school show a greater variety of subject matter than reading textbooks do, and a greater range of style, text structure and language than the selections of reading material in textbooks. The success of a 'trade' book, as opposed to that of a textbook, depends directly on how well it is liked by the

children who read it. A tradebook will remain in print and continue to be read by large numbers of children if it has literary qualities which are perceived and liked by its readers. Children identify with characters who are like them in some ways, particularly in being their age or somewhat older. They may also be intrigued by a particular kind of story or amused by the imaginative use of characters and the expressive qualities of language--puns, jokes, exaggerations and so on. Older children understand generalizations and causal relations better than very young school children. None of these qualities of a book could be easily measured by a readability formula in a way which would distinguish between books which are likely to appeal to children of a particular age and those which probably will not.

In dealing with trade books for children, the best means of matching children of a particular age and reading ability with books they will like is not by formula, but by the judgment of a person who knows children and books. Although there has been mixed success in using people to judge the difficulty of books (Klare, 1984), it would seem unlikely on the face of things that readability formulas could do any better, at least with tradebooks. What sets trade books apart is that they are generally <u>not</u> edited in accordance with readability levels, as textbooks generally are. This is true also of some very popular children's periodicals on science and current events.

The features of tradebooks which make them popular are those which formulas are not sensitive to. In fact, many textbooks make use of excerpts from previously successful tradebooks, which are often better written than selections created specifically for reading textbooks. It is interesting that a current research project on basal readers in primary grades shows that there are fewer discontinuities and unclear references to antecedents in stories excerpted from trade books than in stories written for basal readers (L. Meyer, p.c.).

Further, the use of persons to judge the approximate level of a trade book makes use of already available resources. Librarians in school and local libraries have direct experience with which books get read and by how many children. They are also often asked to suggest books to children of particular age levels and reading ability, with a certain amount of feedback of how well their suggestions were received. There are also people who read all the trade books published in order to review them in publications which in turn are used to advise librarians in buying new books. They have some confirmation of their judgment of the quality and age level of a book in the subsequent success or failure of the book. So librarians and reviewers of children's books have a great deal of first-hand contact with a large number of books and with successive populations of children. They also have continuing feed-back, from the children and from sales figures, of how accurate their judgment is. This

judgment is based on a number of factors and on sensitivity not only to specific features but also to their interactions in a particular book. This experience and ability to make judgments can be used as a substitute for formulas, provided one avoids unrealistic expectations; estimates are approximate and fall within broad age levels, such as grades 3 to 6, varying also with reading ability. Readability formulas are probably not any more accurate, given that the reading levels given by a particular formula may be in error by one or two grade levels.[1]

Languages Other than English

As Klare has noted in his surveys of research on readability formulas, there have been attempts to extend readability formulas to languages other than English. The languages in question have usually been European languages whose syntax and word structure are not very different from English. They are also languages with extensive written literature, both for adults and children. As various countries and language communities within countries attempt to find textbook material suitable for different levels of schooling and reading ability, it is possible to assess the relative merits of the formula style of approach and the alternatives which make use of existing resources.

Language unlike English in structure, writing system, etc. Although English is one of the two national languages of India, there are also a large number of regional languages used in different states. For example, Marathi is the majority language

of the state of Maharashtra, but the Kannada language, of a different family, is used by a substantial minority. Both languages like Marathi and Kannada are very different in syntax from English, and have much more complex morphology, so that the structure of words is quite different from English. There are long literary traditions in most of the languages of India, but they are primarily concerned with religion and classical themes. Much is written in an archaic or literary style far removed from the contemporary spoken languages. The writing systems are generally based on the syllable, except for Urdu, which is written in the Perso-Arabic script which may omit vowels. In either case, it is not clear what counting 'letters' would mean as an index of word complexity.

A current educational project now going on in India is to create tests of reading achievement in seven of the regional languages. To do this, and to create reading materials for particular grades, it is necessary to have some idea of which texts are generally within the reading ability of children at a particular grade level. No official norms currently exist; in fact, one of the goals of creating the tests of reading achievement is to establish some norms for state educational bodies. There were several ways of approaching this task. One would have been to take the readability tradition used in the U.S., and to apply it to the seven regional languages with modifications in the sampling procedure--counting syllables or

characters, and to the approximate grade levels, as established by samples of texts read by groups of children. This approach in effect makes the creation of the means of assessing texts into the goal of establishing norms, at great expense of time and effort. The alternative which has actually been taken has been to find a group of texts known by experience to be appropriate for the age level taking the text, which is 12 years. These texts were chosen by teachers who have had experience with that level of development and school achievement in students. The texts and the measures of comprehension, which are comprehension questions, are being tried out on samples of students, and those that give the most consistent responses will be used in the test of reading achievement.[2]

This approach makes use of information which is already available, the experience of teachers, and applies it directly to the creation of the test, which is the primary goal. As long as there is a pool of teachers who teach reading in a particular language, it will be possible to draw again upon the judgments of teachers to create new versions of the test. This reliance on the judgment of experienced and intelligent people has probably saved a number of years which would otherwise have been spent in recalibrating readability formulas. It directly addresses the educational goal of finding out the norms for reading achievement. It appears to be a wise use of time, human

resources, and money well suited to the circumstances in which the project is being done.

Languages without previous traditions of writing. In the previous example, the basis for a test of reading ability was an educational tradition which already exists for the languages in question. School primers and other reading material have been in use for a number of years, giving the teachers some first hand knowledge of the problems children have in reading. If there is no currently existing stock of texts used for teaching reading, it is difficult to know how to create texts for teaching reading that present written language in the right order of increasing difficulty. This is the problem faced by the Yupik community of Alaska, who want to try to preserve their language (along with English) by teaching their children to learn to read with Yupik as the medium. Needless to say, the sentence and word structure of this language are very different from English. Without such intervention, the language will soon be lost as children learn only English, from television and movies, as well as school. In this situation, it would not be a good method of teaching reading to use text materials which are too hard for the children, or which are too simple and not appropriate for older children.

Instead of trying to adapt English-based readability formulas to Yupik, the members of the community have tried to draw on their own knowledge and experience as speakers of Yupik. One of the approaches being tried out is to study the stylistic

features of spoken Yupik, to record and analyze how people give information, tell stories, and explain procedures to children. These oral texts, and the general features of style, can then be transferred to the written medium and tried out on groups of children. In this way texts in Yupik can be created for some different age groups, though not necessarily graded into very fine grade level distinctions.[3]

In all of the above examples, consideration is given to exactly what features of the situation, texts or readers that would make the use of readability formulas inappropriate for grading or simplifying texts. In place of formulas, a close analysis of the features of the text, readers or situation allowed existing resources to be used instead. In some cases, the alternative to formulas is deliberately chosen over formulas. But in other cases, there really is no choice--formulas could not be used without radical alteration requiring years of research. The results are not known in all cases, and when they are known, they may not be startling. All that has been shown is that some success can be obtained by paying attention to actual readers, texts and features of language. But the question is not whether alternatives to readability formulas are significantly more successful than the use of readability formulas. Sometimes they are, sometimes not. But each attempt to deal with non-abstract properties of texts and readers adds to the general sum of knowledge about how language is understood.

Research on Language Processing

In the previous section, it was pointed out that there are often cheaper, more direct methods of increasing or predicting comprehension of written materials which make use of already existing resources--the experience of teachers, the knowledge which readers are able to bring to the reading of texts. In this section, in contrast, research for which the methods are just beginning to be developed is presented. It investigates questions about which very little is known by even the most expert investigators. The properties of readability formulas are contrasted with their implied view of language, with some of the properties of language which we are beginning to have firm evidence for, even if the whole picture of how language is processed is still incomplete. Readability formulas address the issue of what constitutes or reflects complexity of language, or at least this issue may be read into them by implication. Whatever one may feel about the use of readability formulas as applied to educational or technical materials, the issue of what constitutes complexity in language has very great importance in its own right.

Complexity and formulas. Readability formulas typically measure average sentence length, in words or syllables, and word complexity in syllable length or frequency. As has been pointed out innumerable times, these are very superficial linguistic measures, and they were designed to be superficial. They are

superficial because they are easy to define, and are properties of all texts. When other measures are added, they also are ones which are easily defined and counted, such as pronouns of various types. It is often pointed out that these variables are not measures of complexity per se. They have some relation to the factors which actually cause a text to be complex, so that they are really only reflections of the actual causes of complexity. On this view, there is some continuity through a text from the properties of the most superficial aspects of word choice and sentence structure, to syntactic structure and organization of content of words, to the most abstract level of meaning.

But there is not always perfect and continuous correlation of text difficulty and linguistic features. The following passage is difficult to understand:

> Further, the belief about the good that it is good and that about the not good that it is not good are alike and so, too, are the belief about the good that it is not good and that about the not good that it is good. What belief then is contrary to the true belief about the not good that it is not good? Certainly not the one which says that it is bad, for this might sometimes be true at the same time, while a true belief is never contrary to a true one. (There is something not good which is bad, so that it is possible for both to be true at the same time.)
>
> Aristotle, De interpretatione, J. Ackrill trans. p. 67

The length of the sentences alone (3.5 sentences in 100 words) would suggest that the passage is not for elementary school children, but the words are not technical or difficult in themselves, except for contrary. But clearly the meaning of this passage is immensely more difficult to grasp than would be predicted by the language it is written in. This is not to say that the syntax of the passage is simple, or that phrases like the good and the not good are easy to grasp. The meaning is independently more complex than the language it is expressed in, and so the language does not necessarily reflect semantic complexity.

The predictive power of readability formulas rests on a correlation between superficial features and comprehension measured in some way. The surface features are not always assumed to cause difficulties of comprehension. But there is no reason why they should not be sources of difficulty in themselves. Unfamiliar words in written form may be hard to identify and to relate to the reader's mental lexicon. Long sentences may be hard to process simply because there are so many parts to be related to one another. Formulas embody an entirely plausible notion that the capacity of a reader to process a certain amount of information in a given interval can be exceeded, with disruption of comprehension. The problem with this model, which has never been explicitly addressed in research on readability formulas, is that it is completely vague. We

don't know what unit complexity is measured in, whether sentence structure and word properties are measured in the same units, what interval of time they are contained in, whether this is fixed or flexible, or how comprehension is defined.

These are the issues which are surveyed in this section. We can assume that meaning and linguistic expression are not totally dependent one on the other. The research discussed in the last section has shown that complex meaning can be made more understandable by changes of the right kind in surface expression and the way the text is read. It is therefore possible, for some texts, that the language in which they are written itself contributes to the complexity of the text, and impedes comprehension in some way.

Is complexity a fixed value? Certain researchers have recognized that sentence length itself is imperfectly correlated with difficult sentence structure. A long sentence could be long because it consists of a string of coordinate clauses, which present very little problems in processing (1), or because there are subordinate clauses, which are more difficult to process. But not all subordinate clauses are alike, in that internal and left branching clauses (2) are more difficult to process than right branching clauses (3):

(1) A constituent wrote a letter and the letter was informative and the congressman quoted him [the constituent].

(2) The letter [the constituent [the congressman quoted] wrote] was informative.

(3) The letter was informative [which the constituent wrote [who the congressman quoted]].

Researchers such as Botel, Dawkins and Granowsky (1973) addressed this problem directly, in the context of some research being done in linguists and computation. They proposed a parsing program which would assign weightings to internal or embedded structures, like those in (2) and (3), and additional weighting to non-right branching structures, as in (3). In this research program, it was hoped that it would be possible to measure the syntactic density of the sentences in a text at fairly close intervals. Whether such structures are actually more complex to understand as a general class is an empirical issue (cf. Frazier, 1984, p. 184, for evidence which differentiates types of subordinate clauses).

This approach depends on a very general assumption, which is that complexity is a fixed value: if a construction of a particular type is relatively more difficult to understand than a corresponding but different construction, then the complex construction is always complex. This assumption has some intuitive appeal--since the linguistic features which make it complex persist every time the construction is used. That is, if there are perceptual or memory limits which are overloaded by the placement of a subordinate clause in a particular relation, then

this overload should occur whenever the construction occurs. It ought to be possible to use the weighting program or a more taxonomic approach to identify syntactic structures which are complex, provided that complexity is a fixed value. A taxonomy can be based either on a general characterization of syntax or on research on what constructions children acquire after others (Dawkins, 1975), assuming that children learning to deal with sentence structures succeed first with the simple and regular cases and then with the complex and exceptional cases.

There is a great deal of truth to these approaches, except for the fact that complexity seems not to be a fixed value. Complex constructions are not relatively more complex than their counterparts provided that the linguistic context supports the complex construction. What this means is that the complexity of a construction is offset by contextual information which matches the construction.

For example, the research on how children acquire and understand language has always indicated that passive sentences are more complex than active sentences. There seems to be a very plausible explanation for this fact, since passive and other complex sentence types do not indicate grammatical relations of subject and object in the normal way (Davison, 1984). The sentence object in a passive clause is picked out in a way different from the object in an active. If the passive clause is preceded by an antecedent for the object, it takes less

time to understand it than if the sentence is preceded by an antecedent for the agent phrase. This finding is also true of other construction types which are difficult in isolation. The right kind of antecedent in the preceding context shortens the processing time, even though more complex syntactic structure does require more processing time than a less complex structure (Davison & Lutz, 1984).

A syntactic structure which appears to be very complex is the restrictive relative clause (4). It is learned by young children later than other ways of combining sentences, such as coordination (5):

(4) The dog [which ran away from next door] chased our cat.

(5) The dog ran away from next door and it chased our cat.

In some experiments designed to test comprehension in young children (3-6), children often seem to interpret a sentence like (4), with a restrictive clause, as though it had the structure of (5), referring to two separate events both of which are asserted by saying (5). Hamburger and Crain (1982) have proposed that these results do not accurately reflect what young children know about their language. First, Crain and others have found that children as young as three can pick out the correct meaning of sentences like (4) when they are asked to point to pictures instead of making dolls act out situations, which is a more complex task. Second, four-year old children were able to produce and understand restrictive relative clause constructions

correctly when the relative clauses were used appropriately, with the right context. The context must contain various assumptions--that the event described by the relative clause has already occurred and is known to the speaker and hearer, that there is something which is being described by the relative clause, and the information in the relative clause helps to pick what that referent is. The clause which ran away from next door in (4), helps to distinguish a particular dog from all the other dogs in the discourse context, and is not used just as a way of describing a dog, as it is in (5). Restrictive relative clauses are more complex only if used in isolation without appropriate support from the situation in which they occur.

This conclusion should have been obvious, since language is used for communication. The grammar of a language contains many forms for expressing meaning, some more complex than others. The more complex forms are not gratuitous, not just ways of communicating in more enigmatic and difficult ways. They are instead exploited for expressing complex combinations of grammatical, semantic and contextual information in very efficient ways. Hence, complexity is a feature of syntactic structures, but it is relative and not absolute. If complex structures are tested in their appropriate environments, they turn out to be less complex than in isolation. There is some tradeoff between inherent complexity and efficiency of communication.

How is complexity measured in experimental situations? Earlier research on how syntactic structures are comprehended gave very discouraging results. There seemed to be no effects, or very weak effects, of varying syntactic structures. It appeared that syntactic structure did not enter into comprehension in any interesting way, even when children were the subjects, and if anyone should have problems with understanding complex structures, it should be children in the age range before grammar is fully learned. But an explanation has emerged from in the last ten years or so.

The problem is in using memory as a test for the processing of syntactic structures. Memory (recall, recognition) is relevant for testing comprehension of information, the content of sentences. But as studies like Bransford, Barclay and Franks (1972) showed, people have trouble picking out exactly which form of a sentence they have previously read. They recognize sentences which express the meaning of a sentence or group of sentences which were previously read, but are very inaccurate in recognizing exactly the sentences which they saw. The explanation which has been proposed by many researchers is that the surface form of language in a text is not stored in long-term memory in verbatim form. Information is stored in some kind of interpreted form, in which it can be related to previous knowledge, or condensed and used as the basis for inferences (see Johnson-Laird, 1983, for an overview). So all kinds of effects

of syntactic and lexical structure might be found, but not by using long-term memory as the measure.

What should be the way of getting at the effects of syntactic structure and other surface features of language? Language is processed very rapidly. Even when words are repeated back as fast as the subject is able to do that, some kind of interpretation goes on. Marslen-Wilson (1975) showed that subjects can repeat what they have heard within a quarter to a third of a second, and in that time are able to correct or reinterpret small errors in syntax, semantics or sounds. From studies like these, it has been proposed that language processing is rapid, which means that not very much is processed at one time, and it is interactive, which means that many different kinds of information are processed together.

The result has been that research on the effects of syntactic structure in sentence processing has begun to measure what goes on while the sentence is being understood. It appears that the kind of memory used in processing language is short-term or working memory, which takes small chunks of a sentence as what is worked on in short intervals, measured in seconds or fractions of a second. Subjects are asked to respond at certain times by making choices, or producing a word, or simply indicating that they have comprehended a word or a sentence. The time it takes a subject to make a response is measured. More complex tasks of interpretation are assumed to take more time or be more prone to

error and reinterpretation. Research which records eye movements also provides a very exact measurement of how long it took to read sentences with particular sentence structures. For a survey of some current research of this kind, see chapters in Dowty, Karttunen, and Zwicky (1984).

A great deal has been learned from experimental studies like these, as well as from models of how language should be organized, based on what we know about the features of human language and the human cognitive capacity. The picture is far from complete, however, and there is no answer as yet to the question of what makes a text difficult for a given individual to comprehend. These studies do not give information which could be substituted tomorrow for a readability formula. But they do shed light on an issue which is central to language processing and also to readability formulas. That is the nature of short-term or working memory. The idea that one's ability to process language is finite, that only so much can be understood in a given interval, is shared by both readability formulas and research on language processing.

Unfortunately, very little is known about the short-term memory capacity of both adults and children, though it is clear that when this capacity is exceeded, there are difficulties in comprehension. Various factors contribute to overload, including syntactic and semantic density at a given interval, but it is unclear exactly what these factors are, and how they add up

together to being too complex. Individuals differ in how efficient they are at using short-term memory, and children change in the course of development in how efficiently they can use their short-term memory capacity (Case, Kurland, & Goldberg 1982). There is also a tradeoff between capacity and efficiency; the studies which are surveyed in Huggins and Adams (1980) showed that children preferred sentence structures which allowed them to process as much information as possible up to the limits of their capacity to process linguistic information. So it is not clear at present what direct implications this research has for the questions which readability formulas ought to answer but do not. This is a promising area of research, however, in which results should yield a more realistic and useful view of what constitutes complexity in language.

Conclusion

Recent research on readability, in the narrow sense of readability formulas, has concentrated on statistical refinement, computer implementation and greater ease of application. Measurement of other text features than sentence length and word complexity has not been explored, and comparatively little systematic research has been done on how to write texts which are within the range of readers at a given level. The progress which has been achieved has been in the technical area, not in theoretical discussions of what formulas really are representations of, or why they do or do not work. This being

the case, it is unlikely that much progress will be made in the near future in answering some of the real questions which people want answers to--what makes this text difficult for those students to read; how can the text be made better; what texts features are interrelated?

There is now, as in recent years, a certain amount of research on readability in the broader sense, which goes directly to features of texts and readers in specific situations. But unfortunately these studies are not perceived as a systematic and coordinated effort to find an alternative to the formula-like approach. Compared with the predictive power of formulas (which holds for large aggregates of texts and readers and not for smaller groups), the results of a specific attempt to make a text more readable or to match texts and readers may look very small and insignificant. Each such study addresses a fairly small number of factors and since there are so many which might influence the comprehension of a text or a part of it, the results of one study are seldom carried over to further research. Yet there will be no greater understanding of what makes a text complex if research on alternatives to formulas allowed to be demoralized by the comparison of the success in each attempt with the overall predictions of formulas. Certainly in the area of research on the production of texts, it is imperative to understand what goes into the understanding of written language, and to have a model of how comprehension of language works. By

relating research on readability to research in psychology and linguistics on language processing, it is possible to make each attempt to go beyond formulas have some effect. Let us hope that some of the research being done on specific educational and social problems, as well as theoretical research on language processing, will eventually provide the insight into these questions which has eluded us for so long.

References

Beck, I. L., McKeon, M. G., Omanson, R. C., & Pople, M. T. (1984). Improving the comprehensibility of stories: The effects of revisions that improve coherence. Reading Research Quarterly, 19, 263-277.

Botel, M., Dawkins, J., & Granowsky, A. (1973). A syntactic complexity formula. In W. H. MacGinitie (Ed.), Assessment of problems in reading. Newark, DE: International Reading Association, 77-86.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretative approach. Cognitive Psychology, 3, 193-209.

Case, R., Kurland, M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory. Journal of Experimental Psychology, 33, 386-407.

Charrow, R., & Charrow, V. (1979). Making legal language understandable: A psycholinguistic study of jury instructions. Columbia Law Review, 79(7), 1306-1374.

Davison, A. (1984). Sentence processing and the definition of syntactic topic. Language, 60, 797-846.

Davison, A., & Lutz, R. (1984). Measuring syntactic complexity relative to discourse context. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), Natural language parsing: Psychological, computational and theoretical perspectives (pp. 26-66). Cambridge, MA: Cambridge University Press.

Dawkins, J. (1975). Syntax and readability. Newark, DE: International Reading Association.

Dowty, D., Karttunen, L., & Zwicky, A. (1984). Natural language parsing: Psychological, computational and theoretical perspectives. Cambridge, MA: Cambridge University Press.

Flower, L., Hayes, J. R., & Swarts, H. (1980). Revising functional documents: The scenario principle (Tech. Rep. No. 10). Pittsburgh: Document Design Project, Carnegie-Mellon University with American Institutes for Research and Siegel and Gale.

Frazier, L. (1984). Syntactic complexity. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), Natural language parsing: Psychological, computational and theoretical perspectives (129-89). Cambridge, MA: Cambridge University Press.

Green, G. M. & Laff, K. (to appear). Preferences for and comprehension of original and readability-adapted materials. Urbana: University of Illinois, Center for the Study of Reading.

Hamburger, H., & Crain, S. (1982). Relative acquisition. In S. A. Kuczaj (Ed.), Language development (pp. 245-274). Hillsdale, NJ: Erlbaum.

Holland, V. M. (1981). Psycholinguistic alternatives to readability formulas (Tech. Rep. No. 12). Washington, DC: Document Design Project, American Institute for Research with Carnegie Mellon University and Siegel and Gale.

Huggins, A. W. F. & Adams, M. J. (1980). Syntactic aspects of reading comprehension. In W. Brewer, B. C. Bruce, & R. J. Spiro (Eds.), Theoretical issues in reading comprehension (pp. 87-112). Hillsdale, NJ: Erlbaum.

Johnson-Laird, P. N. (1983). Mental models: Towards a cognitive science of language, inference, and consciousness. Cambridge: Harvard University Press.

Klare, G. R. (1963). The measurement of readability. Ames: Iowa State University Press.

Klare, G. R. (1974-75). Assessing readability. Reading Research Quarterly, 10, 62-102.

Klare, G. R. (1984). Readability. In P. D. Pearson (Ed.), Handbook of reading research (pp. 681-744). New York and London: Longman.

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. Science, 189, 226-228.

Omanson, R. C., Beck, I. L., Voss, J. F. & McKeown, M. G. (1984). The effects of reading lessons on comprehension: A processing description. Cognition and Instruction, 1, 45-67.

Raygor, A. L. (1977). Raygor readability estimator hand calculator. Rehoboth, MA: Twin Oaks Publishing Inc.

Rodriguez, N. & Hansen, L. H. (1975). Performance of readability formulas under conditions of restricted ability level and restricted difficulty of materials. Journal of Experimental Education, 44, 8-14.

Walmsley, S., Scott, K. M., & Lehrer, R. (1981). Effects of document simplification in the reading comprehension of the elderly. Journal of Reading Behavior, 13, 237-248.

Footnotes

[1]I am indebted for discussion on these issues to Zena Sutherland, University of Chicago Graduate Library School.

[2]I am grateful to Dr. R. Shreedhar, Central Institute of Indian Languages, Mysore, for information and discussion.

[3]I am indebted to Dr. Anthony Woodbury, Department of Linguistics, University of Texas, for information and discussion.